arXiv:1101.5591v4 [cond-mat.dis-nn] 16 May 2011

Direct, physically-motivated derivation of the contagion condition for spreading processes on generalized random networks

Peter Sheridan Dodds,[1,2,*] Kameron Decker Harris,[1,2,†] and Joshua L. Payne[3,‡]

[1]*Department of Mathematics & Statistics, The University of Vermont, Burlington, VT 05401.*
[2]*Complex Systems Center & the Vermont Advanced Computing Center, The University of Vermont, Burlington, VT 05401.*
[3]*Computational Genetics Laboratory, Dartmouth College, Hanover, NH 03755*

(Dated: May 17, 2011)

For a broad range of single-seed contagion processes acting on generalized random networks, we derive a unifying analytic expression for the possibility of global spreading events in a straightforward, physically intuitive fashion. Our reasoning lays bare a direct mechanical understanding of an archetypal spreading phenomena that is not evident in circuitous extant mathematical approaches.

PACS numbers: 64.60.aq, 89.75.Hc, 87.23.Ge, 05.45.-a, 64.60.Bd

I. INTRODUCTION

Spreading is a universal phenomenon occurring in many disparate systems across all scales, as exemplified by diffusion and wave propagation, nuclear chain reactions, the dynamics of infectious biological diseases and computer viruses, and the social transmission of religious and political beliefs. Many spreading processes take place on networks, or leave a branching network of altered entities in their wake, and over the last decade, studies of contagion on random networks in particular have provided fundamental insights through analytic results for abstract models [1, 2]. Furthermore, in acknowledging the governing roles of the degree distribution [3] and correlations between nodes [1], generalized random networks [4] have been profitably employed in modeling real-world networks [5]. Thus, a clear, physical understanding of the dynamics of contagion processes on generalized random networks provides a crucial analytic cornerstone for the goal of understanding spreading on real-world networks.

Here, we obtain a unifying analytic expression for the possibility of a global spreading event—which we define as the infection of a non-zero fraction of an infinite network—for a broad range of contagion processes acting on generalized random networks and starting from a single infected seed. We provide both a general framework and results for a series of specific random network families, allowing us to reinterpret, integrate, and illuminate previously obtained conditions. Our explanation has obvious pedagogic benefits: While results for these families are known, previous treatments have centered around powerful but non-intuitive and indirect mathematical approaches, typically involving probability generating functions [2, 4, 6]. We show that a global spreading (or cascade) condition can in fact be transparently derived by considering local growth rates of infection only, such that physical contagion processes are manifest in our expressions.

Our derivation readily accommodates networks with an arbitrary mixture of directed and undirected weighted edges, node and edge characteristics, and node-node correlations, and can be extended to other kinds of random networks such as bipartite affiliation graphs [7]. Our argument also applies to contagion processes evolving in continuous or discrete time, and for the latter case, with either synchronous or asynchronous updates. Nodes may also recover or stay infected as the outbreak spreads.

In what follows, we first obtain an inherently physical condition for the possibility of spreading on generalized random networks, and then provide specific treatments for six interrelated classes of random networks.

II. PHYSICALLY-MOTIVATED DERIVATION OF A GENERAL SPREADING CONDITION

Our goal is to intuitively derive a test for the possibility of global spreading from a single seed, given a specific random network and contagion process [8]. To do so, we construct a global spreading condition based on the infection counts of edge-node pairs rather than just nodes. While considering how the number of infected nodes grows is a more obvious and natural framing, and one that has been broadly employed (e.g., the reproduction number in mathematical epidemiology [9]), the growth of 'infected edges' emanating from infected nodes is equally transparent, and opens a door to analytic treatment.

Since generalized random networks, correlated or not, are locally branching networks [4], successful spreading from a single seed must entail nodes becoming infected in response to a single neighbor's infection (such nodes have been termed 'vulnerable' [10]). For any given contagion process, we therefore need only examine the transmission of infection along single edges. Furthermore, suc-

*Electronic address: peter.dodds@uvm.edu
†Electronic address: kameron.harris@uvm.edu
‡Electronic address: joshua.payne@dartmouth.edu

Typeset by REVTEX

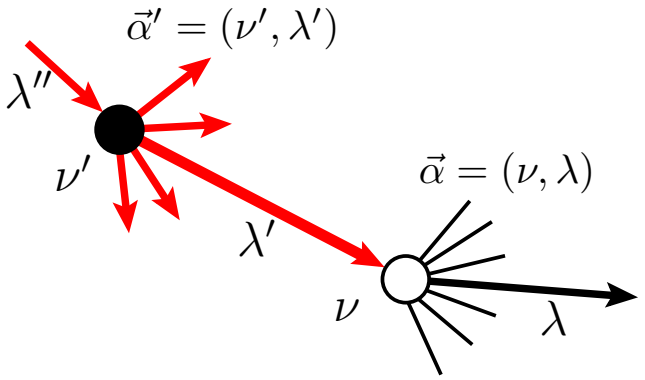

FIG. 1: (Color online) Schematic showing an infection potentially spreading from node-edge pair $\vec{\alpha}' = (\nu', \lambda')$ to node-edge pair $\vec{\alpha} = (\nu, \lambda)$.

cessful spreading leads to exponential growth on random networks when one infected edge, on average, generates more than one new infected edge.

In Fig. 1, we provide a schematic of the spread of a contagious element through a random network. We frame our analysis around the probability that an edge of type λ' 'infects an edge' of type λ through a node of type ν, where by type, we mean individual characteristics such as node or edge age, node degree, edge direction, edge weight, hidden variables, etc. As shown in Fig. 1, a ν' node is already infected due to a λ'' edge and is consequently signalling its infection to its neighbors. In particular, the λ' edge communicates the infection of the ν' node to the ν node and thereby potentially to the marked λ edge. For an infection to spread, we must account for all possible edge-edge transitions incorporating the probability of their occurrence based on (1) network structure and (2) the nature of the spreading process. Our framing leads us to identify node-edge pairs as the key analytic components, as indicated in Fig. 1, and we write $\vec{\alpha} = (\nu, \lambda)$ and $\vec{\alpha}' = (\nu', \lambda')$.

We first consider contagion processes with discrete time updates and one-shot infection chances. By one-shot, we mean that once a node becomes infected, it has one time step to infect its neighbors (excluding the node which infected it), after which no infection can be transmitted. We argue that the growth of the expected number of type $\vec{\alpha}$ node-edge pairs first infected at time t, $f_{\vec{\alpha}}(t)$, follows an exponential growth equation:

$$f_{\vec{\alpha}}(t+1) = \sum_{\vec{\alpha}'} R_{\vec{\alpha}\vec{\alpha}'} f_{\vec{\alpha}'}(t), \tag{1}$$

where $R_{\vec{\alpha}\vec{\alpha}'}$ is what we will call the 'gain ratio matrix', and which possesses a three-part form:

$$R_{\vec{\alpha}\vec{\alpha}'} = P_{\vec{\alpha}\vec{\alpha}'} \bullet k_{\vec{\alpha}\vec{\alpha}'} \bullet B_{\vec{\alpha}\vec{\alpha}'}. \tag{2}$$

The first term $P_{\vec{\alpha}\vec{\alpha}'}$ represents the conditional probability that a type λ' edge emanating from a type ν' node leads to a type ν node. The middle element $k_{\vec{\alpha}\vec{\alpha}'}$ is the number of type λ edges emanating from nodes of type ν, excluding the incident type λ' edge arriving from a type ν' node. The last term $B_{\vec{\alpha}\vec{\alpha}'}$ represents the probability that a type ν node is infected by a single infected type λ' link arriving from a neighboring node of type ν' (the potential recovery of the infected ν' type node is incorporated in $B_{\vec{\alpha}\vec{\alpha}'}$). The first and second elements encode the network's structure, while the third represents the spreading phenomenon, and each term's dependence on $\vec{\alpha}$ and $\vec{\alpha}'$ may be none, part, or whole. In Eq. (2) and below, we use the symbol '$\bullet$' to make clear the composition of the three pieces of the gain ratio matrix.

We can now state the global spreading condition for spreading from a single seed on arbitrarily correlated random networks with discrete time update: the largest eigenvalue of the gain ratio matrix $\mathbf{R} = [R_{\vec{\alpha}\vec{\alpha}'}]$ must exceed unity, i.e.,

$$\sup\{|\mu| : \mu \in \sigma(\mathbf{R})\} > 1 \tag{3}$$

where $\sigma(\cdot)$ indicates eigenvalue spectrum.

Next, we can easily accommodate other types of contagion processes by computing the number of nodes infected a distance d away from the seed rather than as a function of time. The infection probability $B_{\vec{\alpha}\vec{\alpha}'}$ is then computed over all time and is interpreted as the probability that a node of type ν is eventually infected by edge λ'. We now more generally write $f_{\vec{\alpha}}(d+1) = \sum_{\vec{\alpha}'} R_{\vec{\alpha}\vec{\alpha}'} f_{\vec{\alpha}'}(d)$, with $B_{\vec{\alpha}\vec{\alpha}'}$'s role altered and we see that the same global spreading condition arises. Therefore, Eq. (3) applies for contagion processes for which time is continuous or discrete, where nodes may recover, etc., all providing we can sensibly compute $B_{\vec{\alpha}\vec{\alpha}'}$ [11].

III. APPLICATION TO UNDIRECTED, DIRECTED, AND MIXED RANDOM NETWORKS

We now apply our argument to six interrelated classes of random networks, connecting to existing results in the literature. We consider networks with arbitrary degree distributions, mixtures of undirected and directed edges, and node-node correlations based on node degree. Our general global spreading condition takes on specific forms for these networks which are worth deriving individually. We summarize the resulting global spreading conditions in Tab. I.

We generally follow the approach of Boguñá and Serrano [2], who provided a formulation for degree-correlated random networks with mixed undirected and directed edges. We represent nodes by a degree vector $\vec{k} = [\,k_{\mathrm{u}}\ k_{\mathrm{i}}\ k_{\mathrm{o}}\,]^{\mathrm{T}}$ where the entries are, respectively, the number of undirected (or bidirectional) edges between a node and its neighboring nodes; the number of directed edges leading in to a node; and the number of directed edges leading away from a node. For random networks, the explicit inclusion of undirected edges is necessary for modeling instances of mutual influence between nodes,

Network:	Local Growth Equation:	Gain Ratio Matrix:
I. Undirected, Uncorrelated	$f(d+1) = Rf(d)$	$R = \sum_{k_\mathrm{u}} P^{(\mathrm{u})}(k_\mathrm{u} \mid *) \bullet (k_\mathrm{u} - 1) \bullet B_{k_\mathrm{u},*}$
II. Directed, Uncorrelated	$f(d+1) = Rf(d)$	$R = \sum_{k_\mathrm{i},k_\mathrm{o}} P^{(\mathrm{i})}(k_\mathrm{i}, k_\mathrm{o} \mid *) \bullet k_\mathrm{o} \bullet B_{k_\mathrm{i},*}$
III. Mixed Directed and Undirected, Uncorrelated	$\begin{bmatrix} f^{(\mathrm{u})}(d+1) \\ f^{(\mathrm{o})}(d+1) \end{bmatrix} = \mathbf{R} \begin{bmatrix} f^{(\mathrm{u})}(d) \\ f^{(\mathrm{o})}(d) \end{bmatrix}$	$\mathbf{R} = \sum_{\vec{k}} \begin{bmatrix} P^{(\mathrm{u})}(\vec{k} \mid *) \bullet (k_\mathrm{u} - 1) & P^{(\mathrm{i})}(\vec{k} \mid *) \bullet k_\mathrm{u} \\ P^{(\mathrm{u})}(\vec{k} \mid *) \bullet k_\mathrm{o} & P^{(\mathrm{i})}(\vec{k} \mid *) \bullet k_\mathrm{o} \end{bmatrix} \bullet B_{k_\mathrm{u}k_\mathrm{i},*}$
IV. Undirected, Correlated	$f_{k_\mathrm{u}}(d+1) = \sum_{k'_\mathrm{u}} R_{k_\mathrm{u}k'_\mathrm{u}} f_{k'_\mathrm{u}}(d)$	$R_{k_\mathrm{u}k'_\mathrm{u}} = P^{(\mathrm{u})}(k_\mathrm{u} \mid k'_\mathrm{u}) \bullet (k_\mathrm{u} - 1) \bullet B_{k_\mathrm{u}k'_\mathrm{u}}$
V. Directed, Correlated	$f_{k_\mathrm{i}k_\mathrm{o}}(d+1) = \sum_{k'_\mathrm{i},k'_\mathrm{o}} R_{k_\mathrm{i}k_\mathrm{o}k'_\mathrm{i}k'_\mathrm{o}} f_{k'_\mathrm{i}k'_\mathrm{o}}(d)$	$R_{k_\mathrm{i}k_\mathrm{o}k'_\mathrm{i}k'_\mathrm{o}} = P^{(\mathrm{i})}(k_\mathrm{i}, k_\mathrm{o} \mid k'_\mathrm{i}, k'_\mathrm{o}) \bullet k_\mathrm{o} \bullet B_{k_\mathrm{i}k_\mathrm{o}k'_\mathrm{i}k'_\mathrm{o}}$
VI. Mixed Directed and Undirected, Correlated	$\begin{bmatrix} f^{(\mathrm{u})}_{\vec{k}}(d+1) \\ f^{(\mathrm{o})}_{\vec{k}}(d+1) \end{bmatrix} = \sum_{k'} \mathbf{R}_{\vec{k}\vec{k}'} \begin{bmatrix} f^{(\mathrm{u})}_{\vec{k}'}(d) \\ f^{(\mathrm{o})}_{\vec{k}'}(d) \end{bmatrix}$	$\mathbf{R}_{\vec{k}\vec{k}'} = \begin{bmatrix} P^{(\mathrm{u})}(\vec{k} \mid \vec{k}') \bullet (k_\mathrm{u} - 1) & P^{(\mathrm{i})}(\vec{k} \mid \vec{k}') \bullet k_\mathrm{u} \\ P^{(\mathrm{u})}(\vec{k} \mid \vec{k}') \bullet k_\mathrm{o} & P^{(\mathrm{i})}(\vec{k} \mid \vec{k}') \bullet k_\mathrm{o} \end{bmatrix} \bullet B_{\vec{k}\vec{k}'}$

TABLE I: Summary of local growth equations and ratios for six classes of random networks and general contagion processes. These equations describe the expected early growth in infection counts, represented by f, starting from a single initial infective (or seed). Each gain ratio is written so as to highlight three distinct factors in the following order: (1) the probability of an edge leading to a node of a specific type; (2) the number of infected edges arising from a successful infection; (3) the probability of successful infection of that node. As the forms show and as discussed in the main text, these three factors depend on the nature of the edge potentially transmitting an infection. The gain ratio is a scalar for classes I and II, and a matrix for classes III–VI. When the gain ratio is a scalar, the contagion condition is simply $R > 1$, while for the matrix cases, at least one eigenvalue must exceed 1. Classes I–V are special cases of class VI.

and analytically affords a way of connecting directed networks with undirected ones.

We write the probability that a randomly selected node has degree vector $\vec{k}$ as $P_{\vec{k}}$. We represent correlations between nodes via three transition probabilities: $P^{(\mathrm{u})}(\vec{k}|\vec{k}')$, $P^{(\mathrm{i})}(\vec{k}|\vec{k}')$, and $P^{(\mathrm{o})}(\vec{k}|\vec{k}')$, which are the probabilities of an undirected, incoming, or outgoing edge leading from a vector degree $\vec{k}'$ node to a vector degree $\vec{k}$ node. The superscripts therefore refer to the degree $\vec{k}$ node (these conditional probabilities are defined similarly to those used in [2], but with the directed cases reversed).

As we have argued in general, in finding the global spreading condition for random networks, we have to determine three quantities: (1) the probability that a type λ' edge emanating from an infected type ν' node leads to a type ν node where we may have to condition on $\vec{\alpha}'$ and $\vec{\alpha}$; (2) in the case of successful infection, the resultant number of newly infected outgoing λ type edges emanating from the type ν node; and (3) the probability that the type ν node becomes infected.

We start with the basic case of undirected, uncorrelated random networks with a prescribed degree distribution P_{k_u} (class I). The first of the three quantities is given by the observation that following a randomly chosen edge leads to a degree k_u node with probability $k_\mathrm{u}P_{k_\mathrm{u}}/\langle k_\mathrm{u}\rangle$ [7], which we will write as $P^{(\mathrm{u})}(k_\mathrm{u} \mid *)$ with the '$*$' indicating an absence of correlations. Second, if a degree k_u node is infected, $k_\mathrm{u}-1$ new edges will be infected. And third, we have that a degree k_u node becomes infected with probability $B_{k_\mathrm{u},*}$. Putting these pieces together and summing over all possible values of k_u (since the network is uncorrelated), we arrive at the well known global spreading condition for random networks:

$$R = \sum_{k_\mathrm{u}=0}^{\infty} P^{(\mathrm{u})}(k_\mathrm{u} \mid *) \bullet (k_\mathrm{u} - 1) \bullet B_{k_\mathrm{u},*} > 1. \qquad (4)$$

The local growth equation is simple: $f(d+1) = Rf(d)$. In the case that we set $B_{k_\mathrm{u},*} = 1$, meaning the contagion process is always successful, we have the condition for the presence of a giant component, which was obtained by Molloy and Reed [6] in the alternate form $\sum_{k_\mathrm{u}=0}^{\infty} k_\mathrm{u}(k_\mathrm{u} - 2)P_{k_\mathrm{u}} > 0$. Although Molloy and Reed suggested some intuition for this particular form, we believe the kind of derivation we have provided here is the clearest, most direct formulation. Later, Newman *et al.* [7] arrived at the same result using gener-

ating functions, specifically by examining when the average size of finite components diverged for a family of parametrized random networks, and Watts [10], using the same techniques, obtained Eq. (4) for a random network version of Granovetter's threshold-based model of social contagion [12]. These arguments, while entirely effective and part of a larger exploration of the details of random networks (uncovering, for example, distributions of component sizes), are somewhat opaque and roundabout. Thus, while we could readily rearrange Eq. (4) and our other results below to generate more mathematically clean statements, an essential degree of physical intuition would be lost.

In moving to purely directed networks (class II), we now allow each node to have some number of incoming and outgoing edges, k_i and k_o. The three pieces of the gain ratio R are now: (1) upon choosing a random (directed) edge, the probability the edge leads to a node with degree vector $[k_\mathrm{i}, k_\mathrm{o}]^\mathrm{T}$ is $P^{(\mathrm{i})}(k_\mathrm{i}, k_\mathrm{o} \,|\, *) = k_i P_{k_\mathrm{i} k_\mathrm{o}} / \langle k_\mathrm{i} \rangle$; (2) the consequent number of infected outgoing edges is simply k_o; and (3) the probability of infecting such a node is $B_{k_\mathrm{i},*}$. The global spreading condition for uncorrelated directed networks is therefore

$$R = \sum_{k_\mathrm{i}, k_\mathrm{o}} P^{(\mathrm{i})}(k_\mathrm{i}, k_\mathrm{o} \,|\, *) \bullet k_\mathrm{o} \bullet B_{k_\mathrm{i},*} > 1, \tag{5}$$

and the local growth equation is again $f(d+1) = Rf(d)$. The condition for the existence of a giant component, found by setting $B_{k_\mathrm{i},*} = 1$ as before, was obtained by Newman *et al.* [7], again by determining when the average size of finite components diverges. Newman *et al.*'s version of the condition is an elegant algebraic rearrangement of Eq. (5) as $\langle 2k_\mathrm{o}k_\mathrm{i} - k_\mathrm{o} - k_\mathrm{i} \rangle = 0$; Boguñá and Serrano [2] simplified Eq. (5) further to $\langle k_\mathrm{o}(k_\mathrm{i} - 1) \rangle = 0$ since $\langle k_\mathrm{o} \rangle = \langle k_\mathrm{i} \rangle$. Again, the physics of the process is entirely obscured by these mathematically clean statements.

We next consider random uncorrelated networks with arbitrary mixtures of directed and undirected edges (class III). As shown in Tab. I, the local growth equation now accounts for the expected numbers of undirected and directed edges a distance d from the seed, $f^{(\mathrm{u})}(d)$ and $f^{(\mathrm{o})}(d)$ (outgoing rather than incoming edges are recorded since we have framed our analysis around infected edges leaving infected nodes). In computing the expected values of $f^{(\mathrm{u})}(d+1)$ and $f^{(\mathrm{o})}(d+1)$, we see the gain ratio is a 2×2 matrix built around four possible edge-edge transitions: undirected to undirected, undirected to outgoing, incoming to undirected, and incoming to outgoing. The corresponding components of the gain ratio matrix are $P^{(\mathrm{u})}(\vec{k} \,|\, *) \bullet (k_\mathrm{u} - 1)$, $P^{(\mathrm{u})}(\vec{k} \,|\, *) \bullet k_\mathrm{o}$, $P^{(\mathrm{i})}(\vec{k} \,|\, *) \bullet k_\mathrm{u}$, and $P^{(\mathrm{i})}(\vec{k} \,|\, *) \bullet k_\mathrm{o}$. For all four transitions, the probability of infection is $B_{k_\mathrm{u} k_\mathrm{i},*}$. Summing over all possible degrees, we find global spreading occurs when the largest eigenvalue of the gain rate matrix

$$\mathbf{R} = \sum_{\vec{k}} \begin{bmatrix} P^{(\mathrm{u})}(\vec{k} \,|\, *) \bullet (k_\mathrm{u} - 1) & P^{(\mathrm{i})}(\vec{k} \,|\, *) \bullet k_\mathrm{u} \\ P^{(\mathrm{u})}(\vec{k} \,|\, *) \bullet k_\mathrm{o} & P^{(\mathrm{i})}(\vec{k} \,|\, *) \bullet k_\mathrm{o} \end{bmatrix} \bullet B_{k_\mathrm{u} k_\mathrm{i},*} \tag{6}$$

exceeds unity. The global spreading conditions for pure undirected and directed networks, Eqs. (4) and (5) can be retrieved by setting either $P^{(\mathrm{u})}(\vec{k} \,|\, *)$ and k_u or $P^{(\mathrm{i})}(\vec{k} \,|\, *)$ and k_o equal to zero.

The above three classes of uncorrelated random networks (I: undirected, II: directed, III: mixed) have natural degree-degree correlated versions (IV, V, VI). The derivation of their respective global spreading conditions follows the same argument with two changes. First, averaging over node degrees can no longer be done and the gain ratio matrix now has entries for each possible transition between edge types. Second, all transition probabilities are now properly conditional, e.g., $P^{(\mathrm{u})}(k_\mathrm{u} \,|\, *)$ is replaced with $P^{(\mathrm{u})}(k_\mathrm{u} \,|\, k'_\mathrm{u})$ for pure undirected random networks. Consequently, the gain ratio matrix is a function of the degrees $\vec{k}'$ and $\vec{k}$. The resultant gain ratio matrices and the expanded growth equations agree with expressions obtained by Boguñá and Serrano [2], and are shown in Tab. I

IV. CONCLUDING REMARKS

In summary, we have shown that the possibility of global spreading for contagion processes on generalized random networks can be obtained in a direct, physically motivated fashion. A similar kind of clear approach should apply for finding the probability of global spreading. Our work naturally complements that of Gleeson and Cahalane [13] who solved the fundamental problem of the final size of an outbreak, in a similarly straightforward way for macroscopic seeds and, in the limit, for isolated seeds as well. Obtaining an exact solution for the time evolution of spreading from a single seed remains the last major challenge for these random network models.

Acknowledgments

PSD was supported by NSF CAREER Award # 0846668; JLP was supported by NIH grant # K25-CA134286.

[1] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).
[2] M. Boguñá and M. Ángeles Serrano, Phys. Rev. E **72**, 016106 (2005).
[3] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[4] M. E. J. Newman, SIAM Review **45**, 167 (2003).
[5] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nature Genetics pp. 64–68 (2002).
[6] M. Molloy and B. Reed, Random Structures and Algorithms **6**, 161 (1995).
[7] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[8] If we have a parametrized family of networks and contagion processes then we will be able to identify a phase transition between non-spreading and spreading.
[9] J. D. Murray, *Mathematical Biology* (Springer, New York, 2002), Third ed.
[10] D. J. Watts, Proc. Natl. Acad. Sci. **99**, 5766 (2002).
[11] Our approach does not explicitly require an initial single seed, and indeed randomly distributed isolated seeds possess the same global spreading condition, Eq. (3). If the seeds however constitute a non-zero fraction of the network, then the results of [13] apply.
[12] M. Granovetter, Am. J. Sociol. **83**, 1420 (1978).
[13] J. P. Gleeson and D. J. Cahalane, Phys. Rev. E **75**, 056103 (2007).